

Listed are abstracts from recent papers by IBM authors. Inquiries should be directed to the publications cited.

Automatic Parsing for Content Analysis, F. J. Damerau, *Communications of the ACM* **13**, No. 6, 356-360 (June 1970). Although automatic syntactic and semantic analysis is not yet possible for all of an unrestricted natural language text, some applications, of which content analysis is one, do not have such a stringent coverage requirement. Preliminary studies show that the Harvard Syntactic Analyzer can produce correct and unambiguous identification of the subject and object of certain verbs for approximately half of the relevant occurrences. This provides a degree of coverage for content analysis variables which compares favorably to manual methods, in which only a sample of the total available text is normally processed.

Comparison of a Microprogrammed and a Non-microprogrammed Computer, D. E. Waldecker, *Computer Design* **9**, No. 6, 73-78 (June 1970). Discussed are design flexibility considerations and expected realization of lower LSI cost which influence the computer designer in any decision to use microprogram controls.

On a Deterministic Theory of Estimation and Control, G. W. Johnson, *IEEE Transactions on Automatic Control* **AC-15**, No. 1, 125-126 (February 1970). In a recent paper Johnson outlined certain qualitative equivalences between linear deterministic control systems and linear stochastic optimal control systems. This note analyzes the equivalence properties in more detail and establishes a theorem on the inverse stochastic optimal control problem. This theorem is equivalent to Kalman's circle criterion for the restricted stationary case.

On the Entropy of Context-free Languages, W. Kuich, *Information and Control* **16**, No. 2, 173-200 (April 1970). The information theoretical concept of the entropy (channel capacity) of context-free languages and its relation to the structure generating function is investigated in the first part of this paper. The achieved results are applied to the family of pseudo-linear grammars. In the second part, relations between context-free grammars, infinite labelled digraphs and infinite nonnegative matrices are exhibited. Theorems on the convergence parameter of infinite matrices are proved and applied to the evaluation of the entropy of certain context-free languages. Finally, a stochastic process is associated with any context-free language generated by a deterministic labelled digraph, such that the stochastic process is equivalent to the language in the sense that both have the same entropy.

Abstracts

The Impact of Technology and Organization on Future Computer Systems, J. R. Dailey and H. C. Kuntzleman, *Computer Design* **9**, No. 2, 49-54 (February 1970). A philosophy for future systems design is proposed. Cost performance trade-offs are developed by relating technology specifications and cost to system organization and performance.

Information Pattern for Linear Discrete-Time Models with Stochastic Coefficients, T. P. A. Bohlin, *IEEE Transactions on Automatic Control* **AC-15**, No. 1, 104-106 (February 1970). A linear discrete-time system with constant coefficients has an information pattern that does not grow in complexity with time, an information state. This paper shows that a particular canonical form for such systems, the phase-variable model, retains this property when coefficients vary as Gaussian time series with rational spectra. The distribution of the output conditional on past data is normal, and its parameters, being functions of the information state, can be calculated in real time in a simple way. The property is fundamental for effective optimal control. A priori characteristics for the Gaussian time series must be specified, but a maximum likelihood method is proposed for estimating any unknown characteristics from a long sample of input-output data. Also, a parameter-free statistic is found for testing the validity of the phase-variable model in actual cases.

Optimization of Memory Hierarchies in Multiprogrammed Systems, C. V. Ramamoorthy (University of Texas at Austin) and K. M. Chandy, *Journal of the Association for Computing Machinery* **17**, No. 3, 426-445 (July 1970). The optimization of memory hierarchy involves the selection of types and sizes of memory devices such that the average access time to an information block is a minimum for a particular cost constraint. It is assumed that the frequency of usage of the information is known a priori. In this paper the optimization theory for a single task or program is reviewed, and it is extended to a general case in multiprogramming when a number of tasks are executed concurrently. Another important extension treats the case when memories are available only in indivisible modules. Comparisons with conventional methods of solution as well as computational experience on the multiprogrammed and modular cases are given.

A Real-Time Operating System for Manned Spacecraft, P. W. Weiler, R. S. Kopp, and R. G. Dorman, *IEEE Transactions on Computers* **C-19**, No. 5, 388-398 (May 1970). This paper highlights the salient features of the Real-Time Operating System/360 (RTOS) which is being used as the ground support system for the Apollo space program. RTOS is a basic version of Operating System/360 (OS/360) which has been modified and extended to support the requirements of real-time applications in general, and manned spaceflight in particular. Control program services have been added which allow the use of nonstandard I/O devices and which significantly enhance system performance and responsiveness. In addition to providing those services and facilities required in a real-time environment, RTOS supports all conventional OS/360 services. RTOS includes a multi-jobbing capability.

A Relational Model of Data for Large Shared Data Banks, E. F. Codd, *Communications of the ACM* **13**, No. 6, 377-387 (June 1970). Future users of large data banks must be protected from having to know how the data is organized in the machine (the internal representation). A prompting

service which supplies such information is not a satisfactory solution. Activities of users at terminals and most application programs should remain unaffected when the internal representation of data is changed and even when some aspects of the external representation are changed. Changes in data representation will often be needed as a result of changes in query, update, and report traffic and natural growth in the types of stored information. Existing noninferential, formatted data systems provide users with tree-structured files or slightly more general network models of the data. In Section 1, inadequacies of these models are discussed. A model based on n -ary relations, a normal form for data base relations, and the concept of a universal data sublanguage are introduced. In Section 2, certain operations on relations (other than logical inference) are discussed and applied to the problems of redundancy and consistency in the user's model.

Samplesort: A Sampling Approach to Minimal Storage Tree Sorting, W. D. Frazier and A. C. McKellar (Princeton University), *Journal of the Association for Computing Machinery* **17**, No. 3, 496-507 (July 1970). The methods currently in use and previously proposed for the choice of a root in minimal storage tree sorting are in reality methods for making inefficient statistical estimates of the median of the sequence to be sorted. By making efficient use of the information in a random sample chosen during input of the sequence to be sorted, significant improvements over ordinary minimal storage tree sorting can be made. A procedure is proposed which is a generalization of minimal storage tree sorting and which has the following three properties: (a) There is a significant improvement (over ordinary minimal storage tree sorting) in the expected number of comparisons required to sort the input sequence. (b) The procedure is statistically insensitive to bias in the input sequence. (c) The expected number of comparisons required by the procedure approaches (slowly) the information-theoretic lower bound on the number of comparisons required. The procedure is, therefore, "asymptotically optimal."

Sorting in a Paging Environment, B. S. Brawn, F. G. Gustavson, and E. S. Mankin, *Communications of the ACM* **13**, No. 8, 483-494 (August 1970). This sorting study was part of an extensive measurement project undertaken on the M44/44X, an experimental paging system which was conceived and implemented at IBM Research in order to explore the virtual machine concept. The study was concerned with the implementation of sorting procedures in the context of the dynamic paging environment characteristic of virtual memory machines. Descriptions of the experimental sort programs and analysis of the performance measurement results obtained for them are presented. The insight gained from the experimental effort is used to arrive at a set of broad guidelines for writing sort programs for a paging environment.

The Synthesis of Redundant Threshold Logic Elements, R. E. Lyons, *IEEE Transactions on Computers* **C-19**, No. 5, 429-443 (May 1970). This paper is devoted to the development of procedures for synthesizing threshold-logic elements which can be employed in redundant networks based on the von Neumann multiplexing principle. A major result is the statement of rules for synthesizing elements whose outputs are correct despite the presence of e or fewer errors on the input lines. Several theorems are proved and used as the basis for establishing synthesis rules for four canonical redundant threshold element specifications. Some of these canonical elements are optimum in the sense that they require the minimum sum of weights or minimum sum of input bundle sizes.